DOCUMENT RESUME

ED 316 054 FL 018 377

AUTHOR Cheng, Chin-Chuan
TITLE Quantifying Affinity among Chinese Dialects.
INSTITUTION Illinois Univ., Urbana. Language Learning Lab.
REPORT NO LLL-T-10-86
PUB DATE Aug 86
NOTE 25p.
AVAILABLE FROM Language Learning Laboratory, University of Illinois at Urbana-Champaign, G70 Foreign Languages Building, 707 South Mathews, Urbana, IL 60801.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Statistical Data (110)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Chinese; Classification; Comparative Analysis; *Contrastive Linguistics; *Dialects; Mathematical Applications; *Phonology; Research Methodology; Statistical Analysis; *Structural Analysis (Linguistics); Tone Languages; Uncommonly Taught Languages; Vocabulary

ABSTRACT

A study of the relationships between Chinese dialects based on a quantitative measure of dialect affinity is summarized. First, tone values in all the dialect localities available in the early 1970s were used to calculate the dialectal differences in terms of tone height with respect to the "yin and yang" split. In the late 1970s, calculations of lexical correlations among Chinese dialects were performed and the degree of affinity was suggested as a way of viewing dialect relations. More recently, the correlation method was applied to phonological elements for measurements of relatedness. In this report, the earlier attempts at quantification are reviewed and the latest results are presented. Seventeen dialects are included in the study. The quantitative results are found to support what is already known about the qualitative relationships of the dialects. The differential roles played by lexicon, initials, finals, tones, and some combinations in enlarging or compressing distances between dialects are highlighted. The most significant contribution of the quantitative study is seen to be the establishment of a methodology for calculating dialect closeness and determining dialect subgrouping. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED316054

Language Learning Laboratory

LLL

LLL-T-10-86
AUGUST 1986

The School of Humanities

University of Illinois
at Urbana-Champaign

QUANTIFYING AFFINITY AMONG CHINESE DIALECTS

CHIN-CHUAN CHENG

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Cheng, C. C.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

LANGUAGE LEARNING LABORATORY
School of Humanities
University of Illinois at Urbana-Champaign

Technical Report No. LLL-T-10-86

QUANTIFYING AFFINITY AMONG CHINESE DIALECTS

Chin-Chuan Cheng

Director of Language Learning Laboratory
Professor of Linguistics and of Chinese
University of Illinois, Urbana-Champaign

August 1986

Available from: Language Learning Laboratory, University of Illinois at Urbana-Champaign, G70 Foreign Languages Building, 707 South Mathews, Urbana, Illinois 61801

1. In Search of Quantitative Measurements as Syntheses of Dialect Differences

Linguistic research on dialect differences usually focuses on qualitative aspects of language phenomena.[1] Listings of phonological and syntactic rules and descriptions of certain shared linguistic characteristics among dialects are ensembles of such qualitative knowledge. Chinese dialect subgrouping in the past has been based on comparisons of rules and features. While such knowledge has enabled us in many cases to present convincing arguments for classification of one kind or another, it does not provide a means for measuring the *degree* of closeness among dialects.

In comparing dialects, one may want to count the number of shared rules as a way to measure affinity. The more rules two dialects share, the closer they are in comparison with other dialects. However, one serious problem with such a simplistic tabulation is the lack of differentiation among rules. Some rules are generally considered more important than others in dialect classification. But what makes a rule more crucial than others? There is no straightforward answer to this question. Usually a rule that affects more lexical items is considered to be weightier than one that applies to fewer items. Thus quantitative information is taken into consideration in the process of determining subgrouping. The quantitative knowledge in this type of reasoning, however, is not directly obtained by a rigorous method. It is often derived from a synthesis of scholarly work and is therefore subject to variation from individual to individual. As I see it, one of the objectives of science is to express things in terms of verifiable measurements. Since 1973, I have been attempting to quantify Chinese dialect affinity. First, tone values in all the dialect localities available in the early 1970s were used to calculate the dialectal differences in terms of tone height with respect to the *yin* and *yang* split. In the late 1970s, calculations of lexical correlations among Chinese dialects were performed and degree of affinity was suggested as a way of viewing dialect relations. Recently the correlation method was applied to phonological elements for measurements of relatedness. In this writing, the earlier attempts to quantify dialect differences are summarized and the results of current efforts are presented.

2. Quantitative Aspects of Tones in Chinese Dialects

In the early 1970s, I was particularly intrigued by the statement often made about the *yin* and *yang* tone split due to voiceless and voiced initials. It was often said that the modern *yin* tones, derived from voiceless-initial syllables in Middle Chinese, were higher in pitch than the corresponding *yang* tones, which were derived from voiced-initial syllables. This is a reasonable statement in light of the lowering effect of voiced consonants on pitch. I wanted to know if indeed one could find any numerical tendency to support this statement. Counter-examples were easy to find. For example, in Xi'an the *yin ping* tone value is 21 on a five-point scale of pitch height and *yang ping* is 24 (Beijing University 1964:5). Although this statement was not a rule that covered all dialects without exception, it nonetheless could be a generalization of tendencies. However, tendencies needed to be statistically proven. I therefore collected tonal systems of 737 dialect localities for a quantitative study.

The results of the tabulation of tone heights on the 5-point scale were presented in Cheng (1973a) and are reiterated below:

Ping	2.53
Yin Ping	3.47
Yang Ping	2.83
Shang	3.25
Yin Shang	3.98
Yang Shang	2.75
Qu	2.91
Yin Qu	3.13

[1] An earlier version of this work was presented at the Conference on Languages and Dialects of China, Oakland, California, January 15-19, 1986. I am grateful to the conference participants who made various comments on that version. I would also like to thank Stephen Helmreich for his comments on a later version, many of which have been incorporated in this writng.

Yang Qu	2.46
Ru	3.03
Yin Ru	3.19
Yang Ru	3.29
Zhong Ru	2.50
Xia Ru	2.55

Indeed, with the exception of *ru* tones, the *yin* counterparts are numerically higher than the *yang* tones. This was the first quantitative study that I know of to substantiate the claim that tones derived from the syllables with earlier voiced initials are lower than those derived from voiceless ones.

As the tone file was computerized, it was easy to obtain other types of quantitative information. Some of the numerical arrays presented more precise and concrete descriptions of the dialects, while others showed hints of universals and consequently led us to a different sort of linguistic argumentation. For example, the finding of the average number of tones and the range in the dialects was largely descriptive (Cheng 1973a:96-97):[2]

	mean	range
Northern Mandarin	4.06	3-6
Southwestern Mandarin	4.44	4-7
Xiang	4.91	4-6
Eastern Mandarin	5.33	5-7
Huizhou	6.00	6-6
Kejia	6.10	6-7
Gan	6.33	6-7
Wu	6.85	5-8
Northern Min	7.00	7-7
Southern Min	7.30	7-8
Yue	8.17	6-10

A few single-occurrence cases require some comments. The Northern Mandarin group includes the tones of 347 localities; only one locality has six tones, and this is Wubao in Shanxi as described in Wang Lida (1958). The fact that Bobai of the Yue group has 10 tones is well known. The Hailu version of Kejia spoken in Taoyuan, Taiwan has seven tones (Yang 1957) while the other Kejia localities included in the tabulation all have six tones. This list of the number of tones in Chinese dialects can be updated by adding information from recent publications of dialect surveys such as those of Yang (1984).

This 1973 study and subsequent research using the same tone database produced some results that had implications for understanding of language universals. These implications are discussed in Cheng (1977). I will reiterate two of these points below.

Of all the 3,433 tones in the 737 dialect localities, the number of tones in terms of pitch contours had the following distribution:

Falling	1,125
Level	1,086
Rising	790
Falling-rising	352
Rising-falling	80

Falling tones had the highest occurrence and bidirectional tones were comparatively rare. These findings corroborated views put forth earlier and furnished data for subsequent discussion on naturalness and universals of

[2] The range figures were not given in Cheng (1973). They are now presented to complete this type of information.

tone. For example, the frequency of occurrence of the tone shapes listed above supports Wang's (1967) view that bidirectional (falling-rising and rising-falling) tones are more marked than the other shapes. Gandour (1978) was also able to use the numerical data for his study of the perception of tone.

Comparisons of Chinese dialects in terms of the most frequent tone shape were made with the same tone file. The dialects which had more falling tones than level tones were called falling-tone dominant dialects and those which had more level tones than falling tones were called level-tone dominant dialects as follows:

Falling-tone dominant dialects: Northern Mandarin, Southwestern Mandarin, Huizhou, Gan, Kejia.
Level-tone dominant dialects: Eastern Mandarin, Wu, Xiang, Yue, Northern Min, Southern Min.

But no significant consequence of this classification has ever been discussed. The computerization of the tone file allowed me to make various correlations with the numerical data. Some of the correlations may be inconsequential, while others may become significant in the future.

Correlation and co-occurrence of tones in terms of shape, however, was worth pursuing. The initial finding that if a rising-falling tone occurred in a dialect then a rising tone also occurred told much about the relationships of tones within a system. Subsequently, correlations between different tonal contours were calculated and the following table from Cheng (1977) summarizes such relationships (correlation coefficients ranging from -1 to +1):

	Level	Rising	Falling	Rising-falling
Rising	.8359			
Falling	.8504	.8078		
Rising-falling	-.0953	.1229	-.0531	
Falling-rising	.5386	.4841	.4412	-.0552

Besides showing the quantitative patterns of Chinese dialects, frequencies and correlations can help us evaluate or determine the plausibility of historical reconstructions of tones (see, e.g. Mei 1977).

The tones processed in this quantitative study were all phonetic values on a five-point scale. One naturally wishes to be able to use phonological features in such an analysis. However, phonological features are the results of analysis and are not phonetic data. For example, it is difficult to determine the features of the *shang* (315) and *qu* (51) tones in Beijing with respect to the high-low distinction as their digits range from the lowest 1 to the highest 5. I have argued (Cheng 1973b) that the 315 tone is phonologically a low tone and 51, a high tone. Essentially, the arguments are (a) the 315 tone undergoes tone sandhi before another 315 tone while the 51 tone, like the other obviously high tones (55 and 35), does not undergo such a modification and (b) in English-Chinese bilingual code-switching an unstressed English syllable, whose fundamental frequency is lower than a stressed syllable, often causes a preceding Chinese word in the 315 tone to undergo tone sandhi. This is not the place to give a detailed account of all the arguments for this case. The point, however, is clear: phonological features are determined through thorough analyses of the system. It is not feasible for an individual to do detailed studies of all the 737 local systems before making an attempt to find quantitative patterns. While I would admit that in certain cases the classification of tone height may be mechanical and simplistic, I also feel that numerical studies need to be carried out in order to explore their validity. Chinese dialects cover a large geographical spread of Asia. Surveys in the past have produced much descriptive data for synthesis. This study in tones is but a small beginning. Further quantitative research on this wealth of data should contribute to a better understanding of language in general.

3. Degree of Dialect Closeness

This study in tones was mostly aimed at finding numerical correlates among various tones and among the groups of dialects that have been traditionally established. As I looked at this Chinese dialect subgrouping, I found that a great deal of information was missing. As one drew dialect boundaries, the relatedness between two dialect groups disappeared without any trace. For example, how would one describe the distance between Northern Mandarin and Southern Min? Naturally, one could talk about the sharing, or lack of it, of features. But that only reiterated the classification criteria and thus did not answer the question. One could use isoglosses to trace the

relations. Yet, isoglosses were not indices that could express the degrees of closeness. In 1979 I started looking for a methodology that would yield a numerical index that could be considered as degree of dialect closeness.

In the Sino-Tibetan Linguistics Conference in Paris in 1979, I presented a paper dealing with measurements of dialect distance. Pearson's correlation coefficients in statistics were calculated on the basis of the presence or absence of lexical and phonological forms. The correlation coefficients were then interpreted as indices of dialect closeness. Later, the quantification based on the lexicon was pursued further and the results were reported in Cheng (1982). Phonological measurements, on the other hand, have undergone several revisions.

4. Dialect Affinity Based on Lexicon

The *Hanyu Fangyan Cihui* (Beijing University 1964) was the database for an attempt to quantify dialect affinity in terms of lexical items. Cheng (1982) gives a detailed account of the motivation, procedures, and results of such quantification.

The *Cihui* contains 905 words in Putonghua and their variants in 18 localities. In my computerized file, the variants were arranged in such a way that their occurrences across the 18 dialects were marked as 1 for presence and 0 for absence. Following are some examples of the cases for localities Beijing (designated as a), Jinan (b), Shenyang (c), Xi'an (d), Chengdu (e), Kunming (f), Hefei (g), Yangzhou (h), Suzhou (i), Wenzhou (j), Changsha (k), Nanchang (l), Meixian (m), Guangzhou (n), Yangjiang (o), Xiamen (p), Chaozhou (q), and Fuzhou (r). The code 001A and 001B indicate page 1 (left half) and page 1 (right half) of *Cihui*, where the items are located. The two-digit numbers after "A" and "B" indicate the variants in sequence:

		a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r
太阳	001A01	1	1	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0
日头	001A02	0	0	1	1	0	0	0	0	1	0	1	1	0	0	0	1	1	1
爷	001A03	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
热头	001A04	0	0	0	0	0	0	1	0	0	0	0	0	1	1	1	0	0	0
太阳佛	001A05	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
日	001A06	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
日头公	001A07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
月亮	001B01	1	1	1	1	1	1	1	1	1	0	1	0	0	0	1	0	0	0
亮月子	001B02	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
月光	001B03	0	0	0	0	0	0	0	0	0	1	0	1	1	1	0	0	0	0
月	001B04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
月娘	001B05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0

There were 6,454 variants for the 905 words. As we look at the occurrences of 1 and 0 in the examples, we can see some patterns of association among the dialects. For example, Beijing and Jinan (the first two dialects) have the same occurrences of 1 and 0. That is, these two dialects share the same lexical items. This association is, of course, the basis for making dialect subgrouping in terms of cognates. As we look at a couple of more words, the patterns may be even clearer. However, as we take into consideration a large number of words, the patterns may start to blur. At that point, we cannot rely on our eyes any more; we need some tools to help us find the pattern. This is where quantitative study begins. For example, we can construct a contingency table of the occurrences of "0" and "1" using the data for the first three dialects given above.

		Beijing	
		"0"	"1"
Jinan	"1"	0 (a)	2 (b)
	"0"	10 (c)	0 (d)

		Beijing	
		"0"	"1"
Shenyang	"1"	1 (a)	1 (b)
	"0"	9 (c)	1 (d)

This contingency table summarizes the frequency of occurrence of 0s and 1s for each pair. The correlation of each pair of dialects can be calculated with the *phi* correlation:

$$phi = \frac{bc - ad}{\sqrt{(a + c)(b + d)(a + b)(c + d)}}$$

Applying this correlation equation to the contigency tables for the 12 cases given above we obtained 1.0000 for Beijing-Jinan and 0.8888 for Beijing-Shenyang associations. We thus derived a numerical index for describing dialect distance. The *phi* correlation is the Pearson correlation calculated on nominal-dichotomous data. Thus the tool that I have employed to find associations among the dialects is Pearson's correlation.[3] The correlation coefficients have values ranging from +1 to -1. They indicate both the direction and the strength of relationship between two variables. The sign indicates that the two variables are either positively or negatively correlated. Correlation coefficients for all the pairs of the 18 dialects were calculated.[4] The coefficients are considered as degrees of dialect closeness. Table 1 lists all the coefficients thus obtained.

To graphically represent the relationships, the coefficients were processed with a cluster analysis using unweighted average linking. This cluster analysis first established a group by finding the pair of dialects having the minimum distance. For example, Table 1 shows that the Beijing-Shenyang pair has the highest correlation coefficient (0.6983). Beijing and Shenyang were linked as a group. Then the next closest pair of Beijing-Jinan (0.6715) was found. The sum of the coefficients for the pair Jinan-Beijing (0.6715) and Jinan-Shenyang (0.6421) was taken to derive the average of 0.6568 (1.3136 / 2). Generally speaking, the unweighted average linking method calculates the distance between two clusters as the average distance between all pairs of objects in two groups (see e.g. Zupan 1982). The results of the cluster analysis are given in Figure 1. Thus Figure 1 is a dialect tree showing the grouping and degrees of closeness on the basis of lexical items.

Among the examples of the lexical data listed earlier, we saw that the word 'sun' *ri*, *ritou*, and *taiyang* (in Putonghua reading of the words) were treated with equal weights. We know that *ritou* is *ri* plus the suffix *tou*. Thus *ritou* and *ri* belong to the same word family. On the other hand, *taiyang* is morphologically unrelated to the other words. One therefore wishes to take such a fact into consideration and make appropriate statistical processing. My defense is that a lexical item not belonging to the same word family will not be shared by as many dialects, and therefore the correlation will take into account such a difference.

This quantification approach appears to be similar to the lexico-statistical method called glottochronology (Swadesh 1950) that generated a great deal of excitement during the 1950s. But I feel that the fundamental assumptions are quite different and that the results should be interpreted differently. Glottochronology assumed that there were about 200 "basic" words that could account for genetic relationships. Moreover, it calibrated time-depths of language evolution on the basis that 66% of cognates would be shared by languages 1,000 years after they split off from a parent language. This uniform retention rate of change assumed a linear evolution which has received much criticism.

---

[3] The bivariate data for each pair showing the presence or absence of certain items are considered nominal-dichotomous. The correlation formula for calculating the association between x and y as given in Glass and Stanley (1970) is as follows:

$$r_{xy} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right]\left[\sum_{i=1}^{n}(Y_i - \bar{Y})^2\right]}}$$

[4] The SPSS statistical package on CYBER 175 at the University of Illinois at Urbana-Champaign was first used to calculated the correlation coefficients. Later the BMDP package on CYBER 175 was used for this and the other statistical calculations presented in this paper. The BMDP Cluster Analysis of Variables was written by the Department of Biomathematics, University of California at Los Angeles. It was converted to run on CYBER 175 by BMDP Project, Vogelback Computing Center, Northwestern University.

Our data base, the *Hanyu Fangyan Cihui* is a large collection of common words. The number of cases considered helps us to avoid skewing due to a small sample. As discussed in Cheng (1982), the grouping based on lexical items may reflect both genetic relations and cultural dominance. Words of culturally dominant dialects are often borrowed into other speech communities. The degrees of closeness and the dialect tree can be considered a composite form of genetic history and cultural interactions. Genetic relations will be discussed in the following section.

5. Genetic Affinity Based on Phonology

Genetic relations among languages are usually determined by phonological correspondences, morphological-syntactic features, and lexical cognates. In the past, most treatments of Chinese dialect classification have been concerned with phonological characteristics. For example, Ting (1982) lists 16 features that have been taken into account, and all of them fall within the domain of phonology. In this section, I will present my recent work on dialect affinity in terms of initials, finals, and tones.

In the 1979 paper delivered at the Sino-Tibetan Conference in Paris, I presented correlation coefficients among the 17 dialects given in *Hanyu Fangyin Zihui* (Beijing University 1962). Following the procedures used in preparing the cases for lexical items, I listed the initials in all the dialects and tabulated the frequency of occurrence in each dialect. Correlation coefficients were then calculated. For instance, the number of cases of *p* in the more than 2,700 words were counted in each dialect and listed under the case of initial *p*. Such a procedure, however, may obliterate the historical relations. For example, current reflexes of Middle Chinese *p* (*bang* initial not including *fei*), *ph* (*pang* initial not including *fu*), and *b* (*bing* initial not including *feng*) are as follows:

p:	p, ph, b, m
ph:	p, ph, f, x
b:	p, ph, b, f

In the 1979 treatment, all of the occurrences of *p*, irrespective of historical source, were tabulated under one single case. In order to take into account historical relations, this time I separated the cases for tabulation. The *p* derived from Middle Chinese *p* was considered a case by itself. The *p* from Middle Chinese *ph* was another, and the *p* from Middle Chinese *b* was a third case. Thus the true historical relations were kept. Some examples are listed below. The symbol to the left of the colon delimiter is the Middle Chinese initial and the one to the right is the modern dialect segment. All of the 17 dialects in *Zihui* were processed. This set of 17 dialects is not a subset of the 18 dialects given above in the lexical tabulation. However, the two sets share many common dialects. In the following examples of the frequency of occurrence of the initials, the dialects have a different letter-designation: Beijing (a), Jinan (b), Xi'an (c), Taiyuan (d), Hankou (e), Chengdu (f), Yangzhou (g), Suzhou (h), Wenzhou (i), Changsha (j), Shuangfeng (k), Nanchang (l), Meixian (m), Guangzhou (n), Xiamen (o), Chaozhou (p), Fuzhou (q).

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q
p:p	88	89	88	89	88	87	89	90	89	88	85	86	85	87	85	86	85
p:ph	3	2	3	2	3	5	3	0	0	4	6	6	7	5	7	6	7
p:b	0	0	0	0	0	0	0	2	1	0	1	0	0	0	0	0	0
p:m	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
ph:ph	40	40	38	40	39	39	40	39	38	38	38	40	38	39	37	36	40
ph:p	1	1	3	1	1	1	1	2	3	3	2	1	2	2	4	5	1
ph:f	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
ph:x	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
b:p	38	36	32	38	37	37	38	0	0	72	4	5	10	33	57	39	64
b:ph	39	41	44	38	40	39	38	0	0	5	9	72	67	43	20	37	13
b:b	0	0	0	0	0	0	0	77	77	0	64	0	0	0	0	0	0
b:f	0	0	1	1	0	1	1	0	0	0	0	0	0	1	0	0	0

In preparing the data on initials it was decided to omit alternative pronunciations. The literary readings that contrast with colloquial pronunciations were left out, and in the case of multiple pronunciations, the first one listed in *Zihui*

was chosen. The reason for these omissions was to achieve a uniform total number for each dialect and to ensure correct correspondences among the words in question. The modern reflexes of Middle Chinese initials with frequency of occurrence for the 17 dialects are given in Lu and Cheng (1985). Altogether, there were 470 cases for the tabulation.[5] The correlation coefficients and the dialect tree based on initials are given in Table 2 and Figure 2.

The finals were treated in the same fashion to reflect historical correspondences. For Middle Chinese the unit considered was the final that included the small rime (*yun*), lip rounding (*kai-he*), and division (*deng*). The unit for the modern dialects was the entire final (medial, main vowel, and ending). Some examples are given below. These cases are all reflexes of the *dong* rime, the first rime in *Guang Yun*:

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q
uŋ	42	42	0	45	8	0	0	0	0	0	0	48	48	48	0	0	23
oŋ	0	0	43	0	40	48	0	48	48	48	0	0	0	0	0	28	0
ɔuŋ	0	0	0	0	0	0	48	0	0	0	0	0	0	0	0	0	0
aŋ	1	0	0	0	0	0	0	0	0	0	47	0	0	0	10	19	0
ɔŋ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	36	0	0
øyŋ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14
ɵyŋ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7
∂ŋ	3	4	3	3	0	0	0	0	0	0	1	0	0	0	0	0	0
ouŋ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
u∂ŋ	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
uoŋ	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
iã	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
iaŋ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
ɪŋ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
eŋ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0

Altogether, there were 2,770 cases tabulated. The correlation coefficients and the dialect tree are given in Table 3 and Figure 3.

The tone cases were determined by four Middle Chinese tones and three classes of Middle Chinese initials. The four tones are *ping*, *shang*, *qu*, and *ru* (designated as "1", "2", "3", and "4" below). The three classes of initials are voiceless consonants (designated as "x" below), voiced obstruents and sibilants (designated as "y" below), and sonorants (designated as "z" below). As is well known, the voicing of initials affected the historical development of Middle Chinese tones. For example, the Middle Chinese *ping* tone has split into *yin ping* (if the syllable had a voiceless initial in Middle Chinese) and *yang ping* (if the syllable had a voiced initial) in most dialects. Thus in the examples below, "1x:" designates voiceless initial and *ping* tone in Middle Chinese. Then for each Middle Chinese category the frequency of occurrence of words in tones *ping*, *yin ping*, *yang ping*, etc. (designated as 1, 1A, 1B, etc. below) in the modern dialects was tabulated.[6] Thus historical relations were taken into consideration. These are some examples of the tone cases:

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q
1x:1	0	0	0	515	0	0	0	0	0	0	0	0	0	0	0	0	0
1y:1	0	0	0	412	0	0	0	0	0	0	0	0	0	0	0	0	0
1z:1	0	0	0	185	0	0	0	0	0	0	0	0	0	0	0	0	0
1x:1A	522	511	509	0	512	508	511	536	511	504	537	507	500	504	641	495	538
1y:1A	54	59	55	0	54	56	59	54	50	54	52	51	50	52	81	57	55
1z:1A	6	4	4	0	8	11	10	4	2	7	7	5	8	5	7	6	3

[5] I would like to thank Lu Zhiji for his assistance in tabulating the data and drawing the linkage trees.

[6] The DOC computer file does not make a distinction between the *yin* category of tones and the non-splitting category. For example, *yin ping* (in contrast with *yang ping*) and *ping* (no *yin-yang* split) are both coded as "1 " (the number 1 followed by a space in a two-column slot). Thus in the dialects where there is no *yin-yang* split a single category (such as *ping*) was reconstructed for the computer processing.

1x:1B	1	5	3	0	8	4	4	3	2	5	6	2	9	6	15	7	5
1y:1B	367	354	359	0	357	356	357	376	360	358	383	228	354	361	465	349	372
1z:1B	179	170	176	0	171	172	172	187	174	179	184	2	172	178	223	170	186
1x:2	2	4	5	6	2	2	5	3	0	4	3	5	2	0	4	0	4
1y:2	3	3	2	7	4	3	6	2	0	5	2	3	7	0	6	0	5
1z:2	1	3	3	2	4	3	1	0	0	1	2	2	2	0	7	0	4
1x:2A	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	5	0
1y:2A	0	0	0	0	0	0	0	0	3	0	0	0	0	3	0	6	0
1z:2A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0
1y:2B	0	0	0	0	0	0	0	0	2	0	0	0	0	1	0	5	0
1z:2B	0	0	0	0	0	0	0	0	2	0	0	0	0	1	0	1	0

There were 133 cases altogether. The correlation coefficients and the linked tree are given in Table 4 and Figure 4.

In order to understand the composite effects of initials, finals, and tones on dialect affinity, I combined the data and used the same statistical processing. The correlation coefficients for initials and finals (a total of 3,240 cases) are given in Table 5 and Figure 5. Table 6 shows the composite correlation coefficients for initials, finals, and tones (3,373 cases). Figure 6 shows the composite linked tree.

6. Discussion

The correlation coefficients given the Tables 1 through 6 show the degree of association between each pair of the dialects. Several aspects of dialect relations are discussed in Cheng (1982) and Lu and Cheng (1985). Here in this section I will consider a few points that are of general interest.

I have proposed to use the correlation coefficients thus derived as an indicator of degree of dialect affinity. Statements about dialect distance can now be supported by such numerical measurements. At the same time, earlier statements about dialect closeness can be evaluated in terms of these quantitative findings. For example, Zhan (1981) states that Kejia and Gan are much more closely related than the dialects within the Min group. Our correlation coefficients for the pairs Meixian-Nanchang (representing Kejia and Gan) and Xiamen-Chaozhou (within Min) in terms of lexicon (a), initials (b), finals (c), tones (d), initials and finals (e), amd initials, finals, and tones (f) in that order are as follows:

	a	b	c	d	e	f
Meixian-Nanchang	.2722	.8051	.7823	.6171	.8039	.6869
Xiamen-Chaozhou	.3380	.9726	.4037	.8366	.8573	.8434

The only area where the Meixian-Nanchang pair has a higher correlation coefficient is in the finals. In the other areas and in their combinations, the Min dialects are much closer to each other than Kejia and Gan.

While the coefficients provide a means, never before available, to express numerical degrees of closeness as a characterization of linguistic affinity, the dialect trees largely show what is already established of Chinese dialect subgrouping in terms of qualitative knowledge. This close similarity lends support to the validity of this quantitative methodology.

The dialect trees given in Figures 1 through 6, which were established with cluster analysis, use a uniform plotting scale so that the analyses based on the lexicon, initials, finals, tones and some combinations of them can be meaningfully compared. The analyses show that the lexicon makes the dialects farther apart and that tones, on the other hand, have the most compressing effect, shortening the distances among the dialects. Moreover, the subgroupings established by the trees vary in places. For example, in terms of initials, Beijing and Jinan belong to a subgroup, but in terms of finals, Jinan and Xi'an form a separate subgroup. This discrepancy is not unexpected, since different categories of phonological units may have undergone differing historical evolution. Xi'an and Jinan both have nasalized finals. Beijing and Jinan do not have labiodental affricate initials, which occur in Xi'an. The

subgrouping variations shown in the tree figures may be regarded as reflections of such differences in evolution among phonological units.

A drastic case of variation is shown in the affinity based on tones. Tones contribute to the separation of Taiyuan from Northern Chinese. In terms of initials and finals, considered either individually or in combination, Taiyuan is closely associated with the Northern Dialects. When tones are considered (Figures 4 and 6), it branches out as a separate entity. This dialect in Shanxi, because of its geographical proximity to Northern Mandarin, was often considered one of the Mandarin dialects. Decades ago, however, Forrest (1948) set up a separate group called "Jin" to cover this locality and others in southern Shanxi. Hashimoto (1976-77) did comparative studies of Shanxi dialects and consequently made the term "Jinyu" more visible. Taking into consideration the existence of the entering tone and the weakening and deletion of the nasal endings in this dialect and in others that share the same characteristics, Ting (1982) sets up Jinyu as a branch of the Northern Dialects. Li (1985) on the basis of the existence of the entering tone, also considers Jinyu a separate group distinct from Mandarin Dialects. In our examples of the tone data, we can see that Taiyuan has no *yin-yang* distinction in the even tone, which makes the dialect very different from the others.

7. Implications

As I mentioned at the beginning of this writing, a rule that affects more lexical items is often considered to be weightier than one that applies to fewer items. In this study, the "weight" or "strength" of features is implicitly taken into consideration because the tabulation of association is entirely based on the number of words, in which the feature occurs. While Bloomfield (1933:351) asserts that "In modern terminology, the assumption of sound-change can be stated in the sentence: *Phonemes change*", various investigations into historical changes based on words since Wang (1969) have produced significant results (see e.g. Wang 1977). The frequency of occurrence of certain linguistic units will reflect the "strength" of these units. It is my hope that using words in this type of quantitative study will allow us to make the best synthesis of language affinity.

The most significant contribution of these quantitative studies, in my view, is the establishment of a methodology that provides a means for calculating dialect closeness and determining dialect subgrouping. The methodology can be utilized in measuring and grouping subdialects within a group such as those of Wu in Shanghai and Jiangsu that are currently being studied in great detail.

The discussion above assumes that the coefficients thus obtained are true measurements of degrees of dialect closeness and that the average linking method of cluster analysis is appropriate for this type of data.

As early as 1963, Meyers and Wang (1963) discussed the enormity of the task of selecting the correct relational tree. In this context, it seems that I do not have myriad trees to choose from. These trees are built with cluster analysis. The relations defined by the correlation coefficients already eliminate many logical possibilities. As to which is the *correct* tree, however, there is no straightforward answer. Krishnamurti (1978) and Krishnamurti et al. (1983) using shared innovations have presented another way to establish genetic trees. The notion of shared innovation and the notion of correlation as presented in this paper may eventually yield the same results, but various ramifications have yet to be investigated. Even with the correlation measurements there are a few linkage methods available for tree construction. In this writing we have presented only the results using unweighted average linking. Should other methods be used? What compels us to select one method over others? Can our understanding of language, in turn, contribute to the methodology of numerical taxonomy? These are, of course, our long-term research interests. At present, I feel that the quantification methodology and the results provide us with a way to discuss aspects of language change and relation which were not even conceivable before. On the other hand, scholarship is cumulative, and there is plenty of room for improvement in this type of quantification of dialect affinity.

	Beijing	Jinan	Shenyang	Xi'an	Chengdu	Kunming	Hefei	Yangzhou	Suzhou	Wenzhou	Changsha	Nanchang	Meixian	Guangzhou	Yangjiang	Xiamen	Chaozhou	Fuzhou
Beijing	1.0000																	
Jinan	.6715	1.0000																
Shenyang	.6983	.6421	1.0000															
Xi'an	.6108	.6076	.5881	1.0000														
Chengdu	.4478	.4533	.4254	.4874	1.0000													
Kunming	.4902	.5333	.4818	.5455	.5530	1.0000												
Hefei	.4784	.5008	.4746	.4993	.4802	.5431	1.0000											
Yangzhou	.5110	.5287	.4936	.5396	.5056	.5731	.6014	1.0000										
Suzhou	.2891	.3099	.2866	.3169	.2951	.3547	.3432	.4129	1.0000									
Wenzhou	.2179	.2311	.2104	.2211	.2115	.2492	.2342	.2621	.3128	1.0000								
Changsha	.4613	.4872	.4487	.4836	.4854	.5383	.4836	.5052	.3452	.2610	1.0000							
Nanchang	.4428	.4546	.4179	.4475	.4233	.4767	.4732	.5201	.3755	.2817	.5551	1.0000						
Meixian	.2149	.2123	.1930	.2013	.1658	.1931	.1772	.1912	.1821	.1896	.2260	.2722	1.0000					
Guangzhou	.2401	.2215	.2037	.2090	.1719	.2204	.1993	.2176	.1841	.1949	.2275	.2457	.3022	1.0000				
Yangjiang	.2252	.1984	.1872	.1807	.1480	.1955	.1832	.1942	.1587	.1693	.2008	.2290	.2784	.4776	1.0000			
Xiamen	.1987	.1641	.1428	.1332	.0891	.1248	.1069	.1247	.0798	.1022	.1195	.1331	.1658	.1707	.1860	1.0000		
Chaozhou	.2136	.1737	.1608	.1396	.0984	.1290	.1190	.1300	.0972	.1012	.1353	.1498	.1856	.2118	.2158	.3380	1.0000	
Fuzhou	.2693	.2184	.1920	.2014	.1399	.1613	.1489	.1752	.1230	.1414	.1603	.1844	.1412	.1647	.1568	.2800	.2459	1.0000

Table 1. Correlation Coefficients -- Lexicon

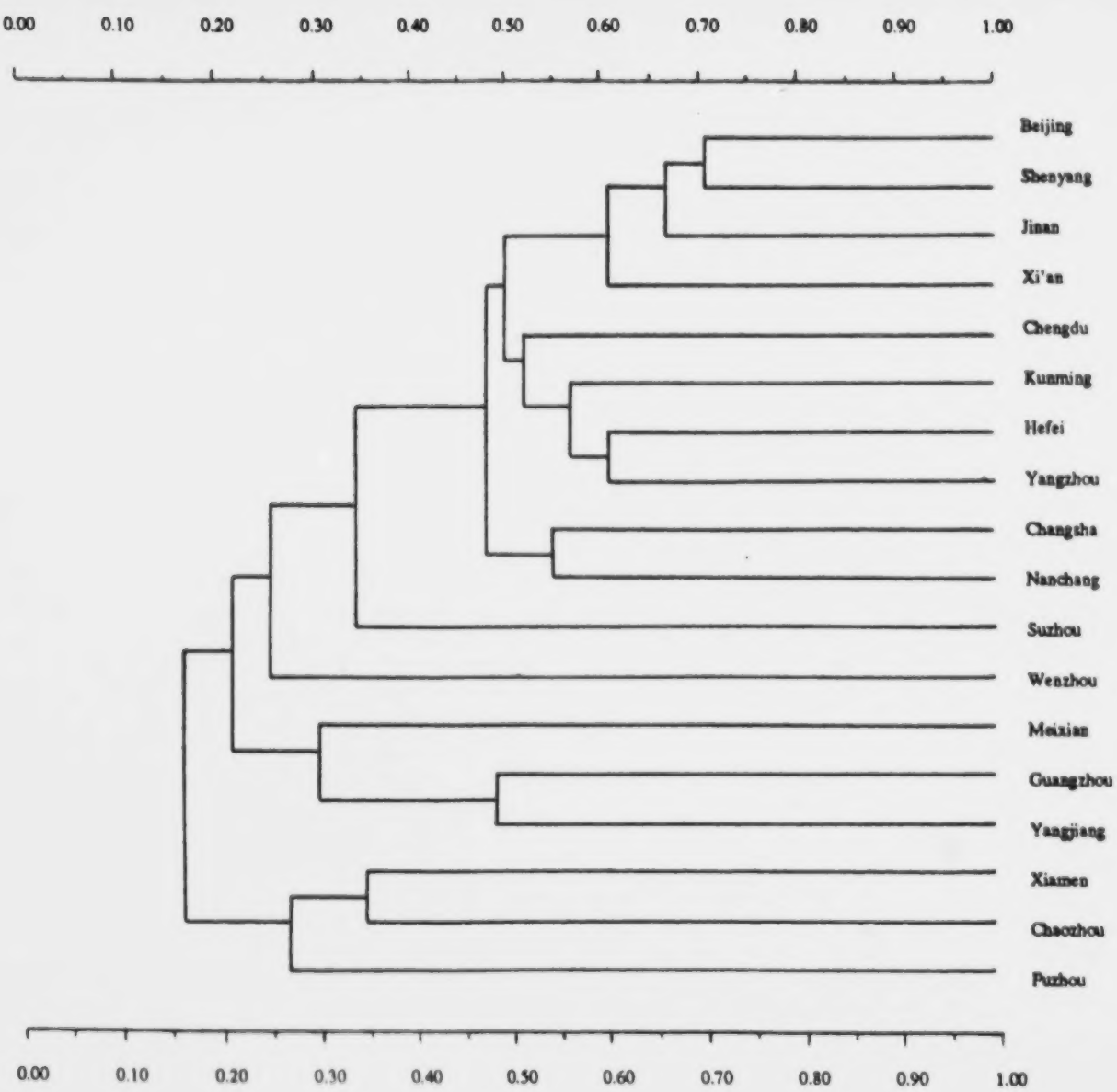

Figure 1. Dialect Affinity Based on Lexicon

	Beijing	Jinan	Xi'an	Taiyuan	Hankou	Chengdu	Yangzhou	Suzhou	Wenzhou	Changsha	Shuangfeng	Nanchang	Meixian	Guangzhou	Xiamen	Chaozhou	Fuzhou
Beijing	1.0000																
Jinan	.9941	1.0000															
Xi'an	.9615	.9628	1.0000														
Taiyuan	.8609	.8610	.9280	1.0000													
Hankou	.7121	.7060	.7592	.8025	1.0000												
Chengdu	.6982	.6967	.7548	.8159	.9858	1.0000											
Yangzhou	.8833	.8779	.9319	.9729	.8183	.8111	1.0000										
Suzhou	.6521	.6492	.7022	.7393	.5828	.5909	.7418	1.0000									
Wenzhou	.6551	.6520	.6996	.7173	.5575	.5557	.7240	.9110	1.0000								
Changsha	.7356	.7307	.7294	.6635	.8548	.8423	.6878	.5712	.5386	1.0000							
Shuangfeng	.5855	.5831	.6024	.5634	.7743	.7573	.5931	.6060	.6330	.7558	1.0000						
Nanchang	.7520	.7608	.8154	.8551	.7021	.7069	.8669	.7330	.7133	.5645	.5762	1.0000					
Meixian	.5806	.5825	.6376	.6958	.5553	.5533	.6959	.6895	.6796	.5198	.4522	.8051	1.0000				
Guangzhou	.5299	.5304	.5501	.5332	.4028	.3826	.5552	.5107	.5302	.4065	.3731	.6029	.7299	1.0000			
Xiamen	.6100	.6038	.6517	.6857	.5753	.5625	.7144	.7091	.6617	.6087	.4944	.7539	.8165	.6592	1.0000		
Chaozhou	.6336	.6301	.6808	.7230	.6054	.5986	.7396	.7441	.6974	.6166	.5250	.8020	.8677	.6935	.9726	1.0000	
Fuzhou	.6863	.6865	.7378	.7677	.6539	.6318	.7950	.7137	.6976	.6715	.5509	.7389	.7808	.6384	.8889	.9017	1.0000

Table 2. Correlation Coefficients -- Initials

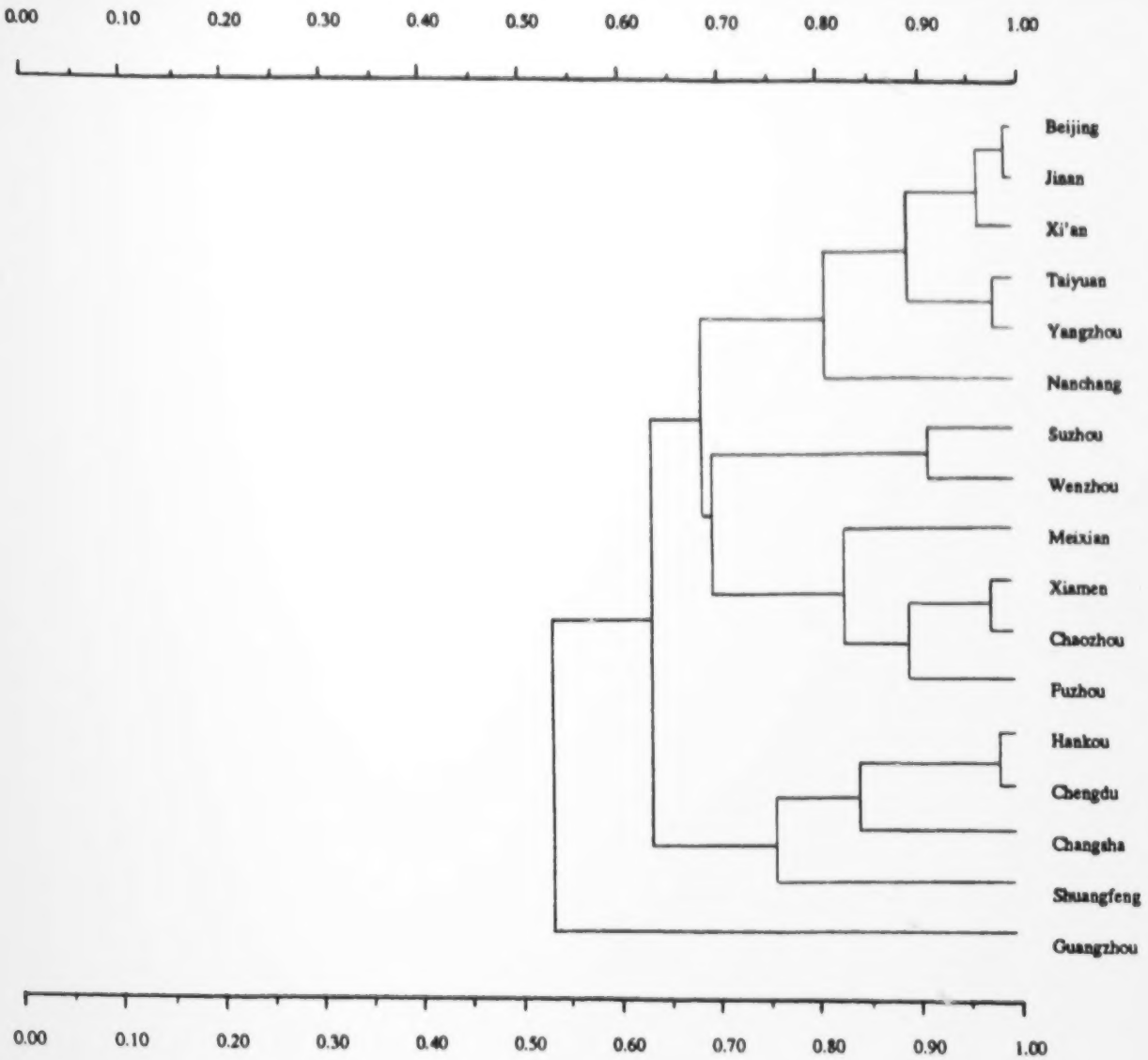

Figure 2. Dialect Affinity Based on Initials

	Beijing	Jinan	Xi'an	Taiyuan	Hankou	Chengdu	Yangzhou	Suzhou	Wenzhou	Changsha	Shuangfeng	Nanchang	Meixian	Guangzhou	Xiamen	Chaozhou	Fuzhou
Beijing	1.0000																
Jinan	.6067	1.0000															
Xi'an	.6823	.7321	1.0000														
Taiyuan	.6246	.4625	.5244	1.0000													
Hankou	.8067	.4312	.6611	.4957	1.0000												
Chengdu	.7615	.3992	.5821	.4533	.8427	1.0000											
Yangzhou	.2605	.3936	.2367	.2621	.2553	.3102	1.0000										
Suzhou	.2005	.1494	.2174	.1056	.3208	.3270	.1475	1.0000									
Wenzhou	.0839	.0885	.1311	.1076	.1404	.1789	.1046	.1593	1.0000								
Changsha	.5036	.2120	.4009	.3649	.6463	.6315	.2231	.2504	.1224	1.0000							
Shuangfeng	.2800	.2579	.2681	.1048	.3144	.3006	.1768	.3173	.0773	.1909	1.0000						
Nanchang	.5018	.2899	.3261	.4446	.4770	.5147	.2771	.1766	.1421	.4460	.1898	1.0000					
Meixian	.4761	.2440	.3199	.4298	.4223	.4647	.2020	.1389	.1277	.3856	.1260	.7823	1.0000				
Guangzhou	.1328	.1221	.1032	.1354	.1194	.0750	.0493	.0275	.0470	.0714	.0397	.2338	.2617	1.0000			
Xiamen	.2781	.1025	.1687	.1537	.2675	.2958	.0973	.0779	.0773	.1936	.1151	.3505	.4300	.1615	1.0000		
Chaozhou	.3559	.2303	.3942	.3167	.4164	.3152	.1679	.1825	.1624	.2833	.1833	.3644	.3600	.1494	.4037	1.0000	
Fuzhou	.2563	.2783	.1905	.2859	.2389	.2253	.1874	.0585	.0532	.1940	.0718	.2742	.2408	.1957	.1167	.2940	1.0000

Table 3. Correlation Coefficients -- Finals

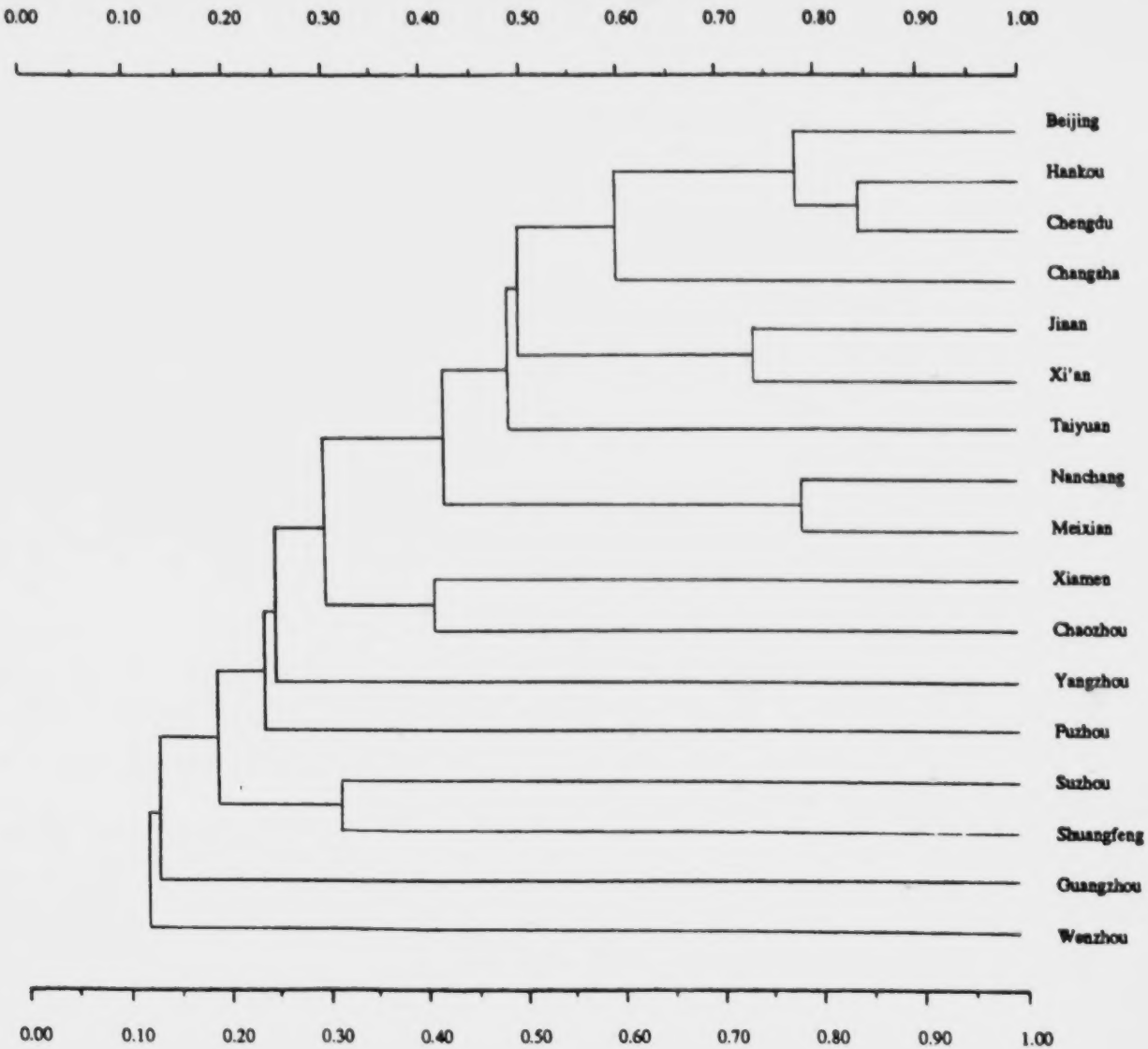

Figure 3. Dialect Affinity Based on Finals

	Beijing	Jinan	Xi'an	Taiyuan	Hankou	Chengdu	Yangzhou	Suzhou	Wenzhou	Changsha	Shuangfeng	Nanchang	Meixian	Guangzhou	Xiamen	Chaozhou	Fuzhou
Beijing	1.0000																
Jinan	.9923	1.0000															
Xi'an	.9832	.9941	1.0000														
Taiyuan	.3028	.2982	.2954	1.0000													
Hankou	.9693	.9497	.9458	.2937	1.0000												
Chengdu	.9670	.9468	.9431	.2932	.9998	1.0000											
Yangzhou	.9405	.9299	.9260	.2990	.9185	.9162	1.0000										
Suzhou	.6844	.6742	.6754	.1127	.6681	.6675	.6633	1.0000									
Wenzhou	.5560	.5491	.5490	-.0041	.5430	.5411	.5357	.8613	1.0000								
Changsha	.6816	.6705	.6728	.0457	.6657	.6649	.7450	.8789	.7484	1.0000							
Shuangfeng	.7187	.6989	.7007	.0464	.7483	.7494	.6792	.9196	.7853	.9044	1.0000						
Nanchang	.6204	.6115	.6125	.0536	.6076	.6070	.6985	.8094	.6614	.8877	.8238	1.0000					
Meixian	.9383	.9283	.9245	.3484	.9168	.9145	.9140	.7634	.6393	.6765	.6966	.6171	1.0000				
Guangzhou	.5618	.5547	.5546	-.0375	.5486	.5466	.5414	.8523	.9800	.7650	.8045	.6779	.6159	1.0000			
Xiamen	.6767	.6661	.6681	.1280	.6611	.6603	.6559	.9920	.8589	.8817	.9181	.8084	.7584	.8469	1.0000		
Chaozhou	.5682	.5613	.5612	-.0068	.5548	.5528	.5475	.8376	.9753	.7418	.7659	.6367	.6525	.9523	.8366	1.0000	
Fuzhou	.6938	.6827	.6848	.1273	.6776	.6768	.6726	.9922	.8554	.8923	.9273	.8198	.7664	.8453	.9985	.8332	1.0000

Table 4. Correlation Coefficients -- Tones

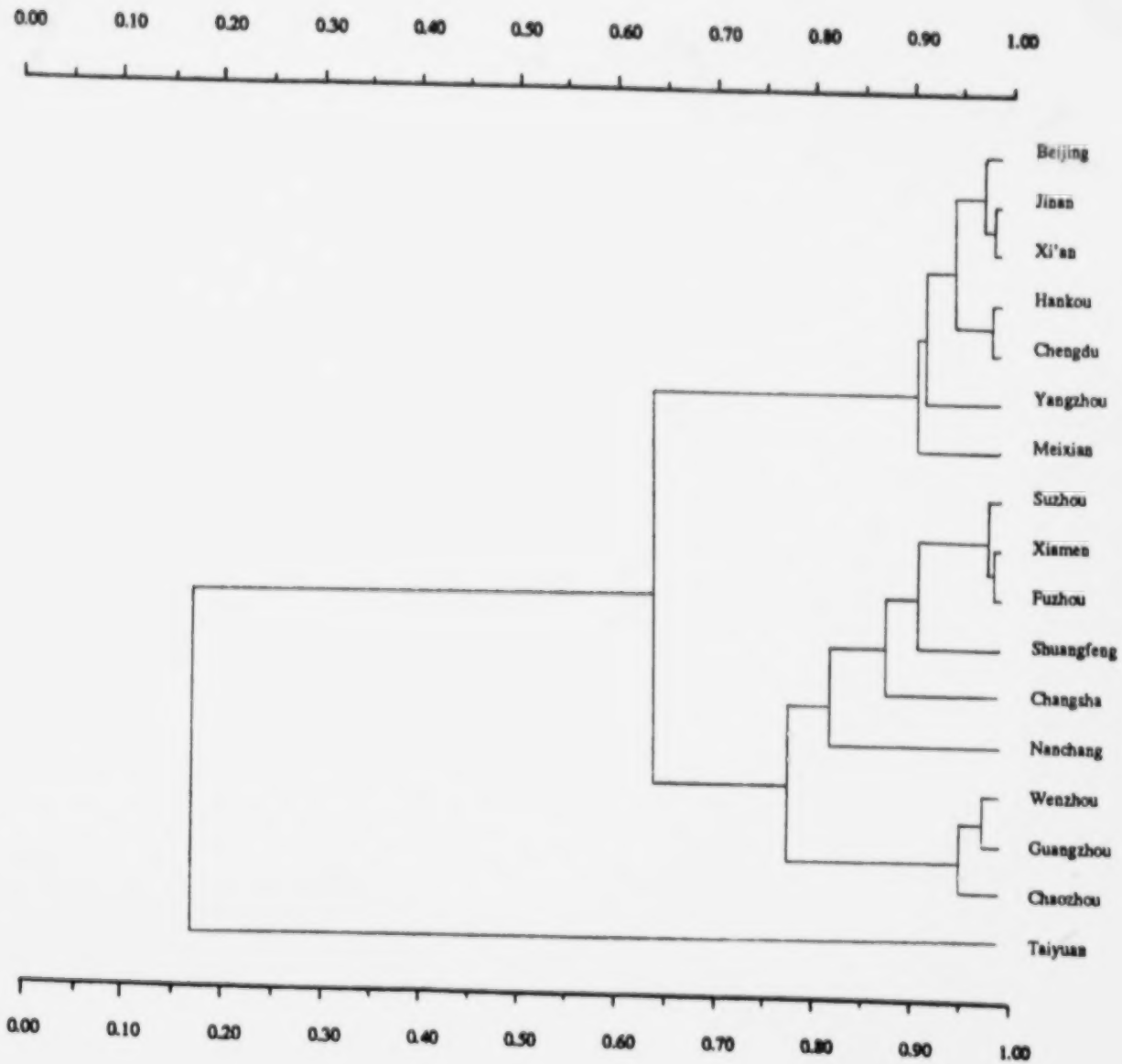

Figure 4. Dialect Affinity Based on Tones

	Beijing	Jinan	Xi'an	Taiyuan	Hankou	Chengdu	Yangzhou	Suzhou	Wenzhou	Changsha	Shuangfeng	Nanchang	Meixian	Guangzhou	Xiamen	Chaozhou	Fuzhou
Beijing	1.0000																
Jinan	.8872	1.0000															
Xi'an	.8847	.8989	1.0000														
Taiyuan	.8003	.7541	.8155	1.0000													
Hankou	.7497	.6407	.7412	.7235	1.0000												
Chengdu	.7278	.6234	.7147	.7183	.9452	1.0000											
Yangzhou	.7163	.7479	.7385	.7741	.6682	.6754	1.0000										
Suzhou	.5493	.5328	.5859	.5831	.5300	.5352	.5980	1.0000									
Wenzhou	.5153	.5128	.5553	.5613	.4614	.4681	.5674	.7260	1.0000								
Changsha	.6823	.5978	.6481	.5933	.8026	.7885	.5718	.5047	.4448	1.0000							
Shuangfeng	.5129	.5042	.5194	.4450	.6478	.6297	.4873	.5409	.4887	.6010	1.0000						
Nanchang	.6916	.6371	.6826	.7437	.6499	.6626	.7060	.5967	.5675	.5480	.4781	1.0000					
Meixian	.5688	.5119	.5689	.6379	.5368	.5448	.5827	.5736	.5568	.5029	.3847	.8039	1.0000				
Guangzhou	.4380	.4344	.4414	.4390	.3459	.3176	.4317	.4038	.4173	.3360	.2990	.5137	.6221	1.0000			
Xiamen	.5429	.4950	.5436	.5645	.5142	.5097	.5736	.5779	.5353	.5217	.4137	.6610	.7391	.5495	1.0000		
Chaozhou	.5800	.5470	.6210	.6331	.5727	.5656	.6120	.6306	.5844	.5507	.4544	.7008	.7654	.5745	.8573	1.0000	
Fuzhou	.5950	.5989	.6133	.6580	.5656	.5433	.6571	.5816	.5600	.5698	.4450	.6340	.6753	.5453	.7362	.7856	1.0000

Table 5. Correlation Coefficients -- Initials and Finals

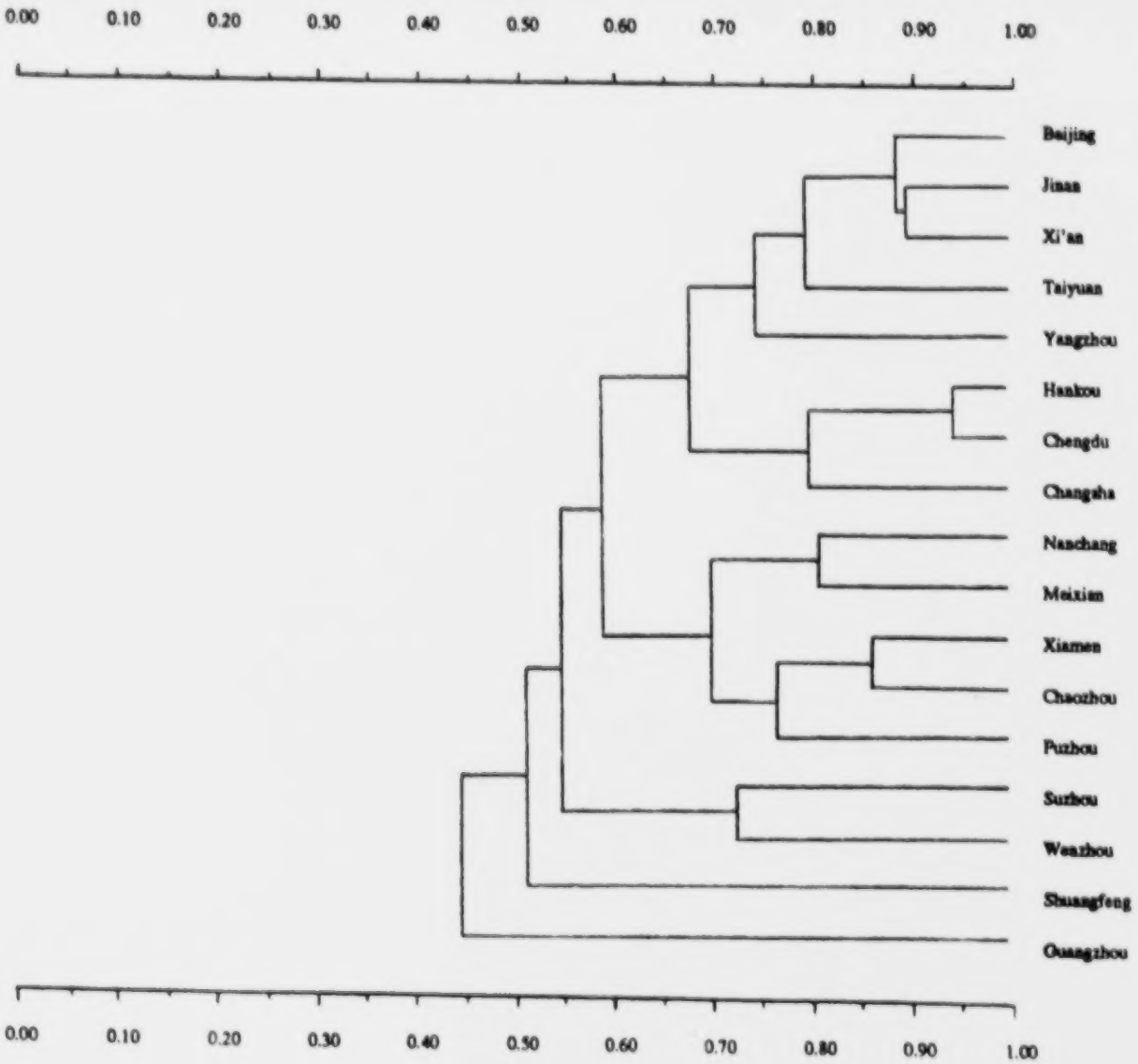

Figure 5. Dialect Affinity Based on Initials and Finals

	Beijing	Jinan	Xi'an	Taiyuan	Hankou	Chengdu	Yangzhou	Suzhou	Wenzhou	Changsha	Shuangfeng	Nanchang	Meixian	Guangzhou	Xiamen	Chaozhou	Fuzhou
Beijing	1.0000																
Jinan	.9668	1.0000															
Xi'an	.9603	.9713	1.0000														
Taiyuan	.4517	.4383	.4474	1.0000													
Hankou	.9176	.8769	.8999	.4261	1.0000												
Chengdu	.9103	.8701	.8913	.4251	.9867	1.0000											
Yangzhou	.8887	.8887	.8848	.4419	.8621	.8618	1.0000										
Suzhou	.6659	.6545	.6691	.2663	.6500	.6506	.6630	1.0000									
Wenzhou	.5671	.5617	.5719	.1755	.5449	.5450	.5652	.8343	1.0000								
Changsha	.6975	.6690	.6831	.2208	.7153	.7115	.7153	.7920	.6865	1.0000							
Shuangfeng	.6836	.6673	.6735	.1823	.7370	.7333	.6495	.8326	.7256	.8375	1.0000						
Nanchang	.6575	.6375	.6491	.2724	.6376	.6406	.7148	.7626	.6537	.8048	.7450	1.0000					
Meixian	.8405	.8180	.8325	.4521	.8180	.8181	.8279	.7221	.6332	.6448	.6286	.6896	1.0000				
Guangzhou	.5488	.5430	.5451	.1309	.5158	.5067	.5330	.7383	.8309	.6621	.6832	.6476	.6355	1.0000			
Xiamen	.6633	.6452	.6583	.2605	.6464	.6445	.6545	.9004	.7955	.8098	.8188	.7809	.7596	.7806	1.0000		
Chaozhou	.5915	.5782	.5969	.2034	.5801	.5770	.5853	.7892	.8718	.7035	.6963	.6734	.7020	.8465	.8434	1.0000	
Fuzhou	.6829	.6762	.6815	.2969	.6643	.6580	.6837	.8865	.7868	.8154	.8118	.7785	.7515	.7698	.9362	.8278	1.0000

Table 6. Correlation Coefficients -- Initials, Finals, and Tones

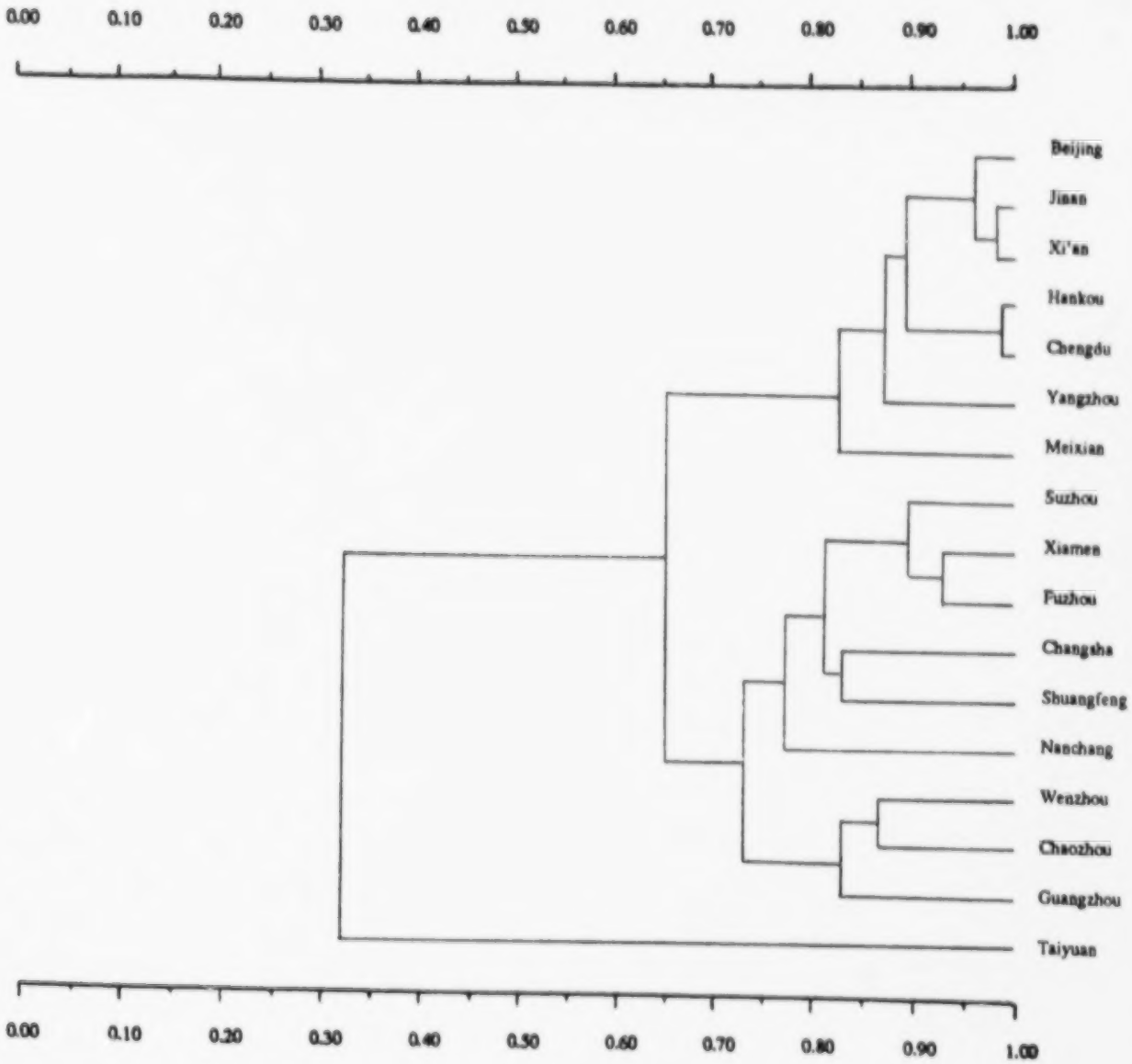

Figure 6. Dialect Affinity Based on Initials, Finals, and Tones

REFERENCES

Beijing University. 1962. *Hanyu Fangyin Zihui*. Beijing: Wenzi Gaige Chubanshe.

_____. 1964. *Hanyu Fanyan Cihui*. Beijing: Wenzi Gaige Chubanshe.

Bloomfield, Leonard. 1933. *Language*. New York: Holt.

Cheng, Chin-Chuan. 1973a. A quantitative study of Chinese tones. *Journal of Chinese Linguistics* 1.93-110.

_____. 1973b. *A Synchronic Phonology of Mandarin Chinese*. The Hague: Mouton.

_____. 1977. Tonal correlations in Chinese dialects: a quantitative study. *Studies in the Linguistic Sciences* 7:2.115-128.

_____. 1982. A quantification of Chinese dialect affinity. *Studies in the Linguistic Sciences* 12.29-47.

Forrest, R. A. D. 1948. *The Chinese Language*. London: Faber and Faber Ltd.

Gandour, Jackson T. 1978. The perception of tone. *Tone: a Linguistic Survey* ed. by Victoria A. Fromkin, 41-76. New York: Academic Press.

Glass, Gene V., and Julian C. Stanley. 1970. *Statistical Methods in Education and Psychology*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Hashimoto, Mantaro. 1976-77. *Comparative studies in Jin dialects*. Asia-Africa Language and Culture Study 12.11-58, 13.77-127, 14.72-132.

Krishnamurti, Bh. 1978. Areal and lexical diffusion of sound change: Evidence from Dravidian. *Language* 54.1-20.

Krishnamurti, Bh.; Lincoln Moses; and Douglas G. Danforth. 1983. Unchanged cognates as a criterion in linguistic subgrouping. *Language* 59.541-568.

Li, Rong. 1985. Guanhua fangyan de fenqu. *Fangyan* 1.2-5.

Lu, Zhiji, and Chin-Chuan Cheng. 1985. Chinese dialect affinity based on syllable initials. *Studies in the Linguistic Sciences* 15:2.127-148.

Mei, Tsu-Lin. 1977. Tones and tone sandhi in 16th century Mandarin. *Journal of Chinese Linguistics* 5.237-260.

Meyers, L. F., and William S-Y. Wang. 1963. *Tree representations in linguistics*. Project on Linguistic Analysis Reports 3, Ohio State University.

Swadesh, Morris. 1950. Salish internal relationships. *International Journal of American Linguistics* 16.157-167.

Ting, Pang-Hsin. 1982. Hanyu fangyan qufen de tiaojian. *Tsing Hua Journal of Chinese Studies* 14.257-272.

Wang, Lida. 1958. Shanxi fangyin zhong de shengdiao yu Putonghua de duiying guanxi. *Fangyan Yu Putonghua Jikan* 5.106-111.

Wang, William S-Y, 1967. Phonological features of tone. *International Journal of American Linguistics* 33:2.93-105.

_____. 1969. Competing changes as a casue of residue. *Language* 45.9-25.

_____. (ed.) 1977. *The lexicon in phonological change*. The Hague: Mouton.

Yang, Shifeng. 1957. *Taiwan Taoyuan Kejia Fangyan*. Taibei: Academia Sinica.

_____. 1984. *Sichuan Fangyan Diaocha Baogao*. Institute of History and Philology Special Publications 82. Taipei: Academia Sinica.

Zhan, Bohui. 1981. *Xiandai Hanyu Fangyan*. Wuhan: Wubei Renmin Chubanshe.

Zupan, Jure. 1982. *Clustering of Large Data Sets*. Chichester: Research Studies Press.